Exhibit S

Motion Tracking: No Silver Bullet, but a Respectable Arsenal

Greg Welch
University of North Carolina at Chapel Hill

Eric Foxlin
InterSense

This article introduces the physical principles underlying the variety of approaches to motion tracking. Although no single technology will work for all purposes, certain methods work quite well for specific applications.

If you read the surveys of motion tracking systems,[1-5] one thing that will immediately strike you is the number of technologies and approaches—a bewildering array of systems operating on entirely different physical principles, exhibiting different performance characteristics, and designed for different purposes. So why does the world need so many different tracking products and research projects to do essentially the same thing?

Just as Brooks argued in his famous article on software engineering[6] that there is no single technique likely to improve software engineering productivity an order of magnitude in a decade, we'll attempt to show why no one tracking technique is likely to emerge to solve the problems of every technology and application.

But this isn't an article of doom and gloom. We'll introduce you to some elegant trackers designed for specific applications, explain the arsenal of physical principles used in trackers, get you started on your way to understanding the other articles in this special issue, and perhaps put you on track to choose the type of system you need for your own computer graphics application. We hope this article will be accessible and interesting to experts and novices alike.

What is motion tracking?

If you work with computer graphics—or watch television, play video games, or go to the movies—you are sure to have seen effects produced using motion tracking. Computer graphics systems use motion trackers for five primary purposes:

- *View control.* Motion trackers can provide position and orientation control of a virtual camera for rendering computer graphics in a head-mounted display (HMD) or on a projection screen. In immersive systems, head trackers provide view control to make the computer graphics scenery simulate a first-person viewpoint, but animations or other nonimmersive applications might use handheld trackers.
- *Navigation.* Tracked devices help a user navigate through a computer graphics virtual world. The user might point a tracked wand to fly in a particular direction; sensors could detect walking-in-place motion for virtual strolling.
- *Object selection or manipulation.* Tracked handheld devices let users grab physical surrogates for virtual objects and manipulate them intuitively. Tracked gloves, acting as virtual surrogates for a user's hands, let the user manipulate virtual objects directly.
- *Instrument tracking.* Tracked tools and instruments let you match virtual computer graphics representations with their physical counterparts—for example, for computer-aided surgery or mechanical assembly.
- *Avatar animation.* Perhaps the most conspicuous and familiar use of trackers has been for generating realistically moving animated characters through full-body motion capture (MoCap) on human actors, animals, and even cars.

No silver bullet

Our experience is that even when presented with motion tracking systems that offer relatively impressive performance under some circumstances, users often long for a system that overcomes the shortcomings related to their particular circumstances. Typical desires are reduced infrastructure, improved robustness, and reduced latency (see the sidebar, "Tracking Latency"). The only thing that would satisfy everyone is a magical device we might call a "tracker-on-a-chip." This ToC would be all of the following:

- *Tiny*—the size of an 8-pin DIP (dual in-line package) or even a transistor;
- *Self-contained*—with no other parts to be mounted in the environment or on the user;

0272-1716/02/$17.00 © 2002 IEEE

Tracking Latency

Have you seen those so-called "gourmet" cookie stands in convenience stores or fast-food restaurants? They usually include a sign that boasts "Made fresh daily!" Unfortunately, while cookie baking might indeed take place daily, the signs don't actually give you the date on which the specific cookies being sold were baked!

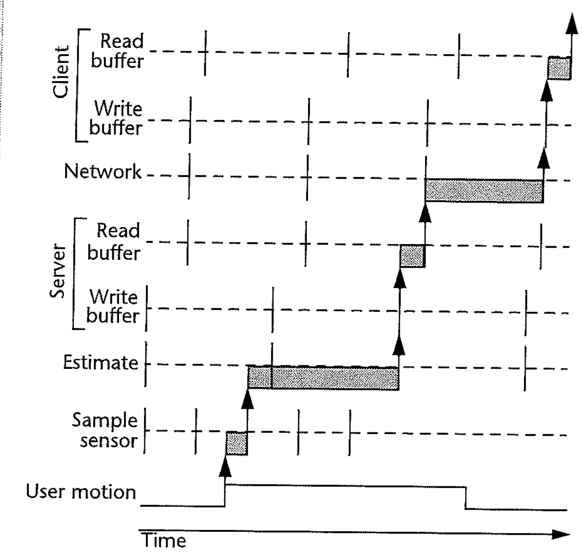
We've found a related common misperception about delay or latency in interactive computer graphics in general, and in tracking in particular. While the inverse of the estimate rate (the period of the estimates) contributes to the latency, it doesn't tell the entire story. Consider our imaginary tracker-on-a-chip. If you send its 1,000-Hz estimates halfway around the world over the Internet, they will arrive at a rate of 1,000 Hz, but quite some time later.

Similarly, within a tracking system, a person moves, the sensors are sampled at some rate, some computation is done on each sample, and eventually estimates pop out of the tracker. To get the entire story, you must consider not only the rate of estimates, but also the length of the pipeline through which the sensor measurements and subsequent pose estimates travel.

As Figure A illustrates, throughout the pipeline there are both fixed latencies, associated with well-defined tasks such as sampling the sensors and executing a function to estimate the pose, and variable latencies, associated with buffer operations, network transfers, and synchronization between well-defined but asynchronous tasks. The variable latencies introduce what's called latency jitter.

Here again there's no silver bullet. In 1995 Azuma showed that motion prediction can help considerably, to a point.[1,2] The most basic approach is to estimate or measure the pose derivatives and to use them to extrapolate forward from the most recent estimate—which is already old by the time you get to see it—to the present time. The problem is that it's difficult to predict what the user will choose (has chosen) to do very far in the future.

Azuma pointed out that the task is like trying to drive a car by looking only in the rear-view mirror. The driver must predict where the road will go, based solely on the view of the past and knowledge of roads in general. The difficulty of this task depends on how fast the car is going and on the shape of the road. If the road is straight and remains so, the task is easy. If the road twists and turns unpredictably, the task is impossible.



A Typical tracker pipeline.

References

1. R. Azuma, *Predictive Tracking for Augmented Reality*, PhD dissertation, tech. report TR95-007, Univ. North Carolina, Chapel Hill, Dept. Computer Science, 1995.
2. R. Azuma and G. Bishop, "A Frequency-Domain Analysis of Head-Motion Prediction," *Proc. Ann. Conf. Computer Graphics and Interactive Techniques* (Proc. Siggraph 95), ACM Press, New York, 1995, pp. 401-408.

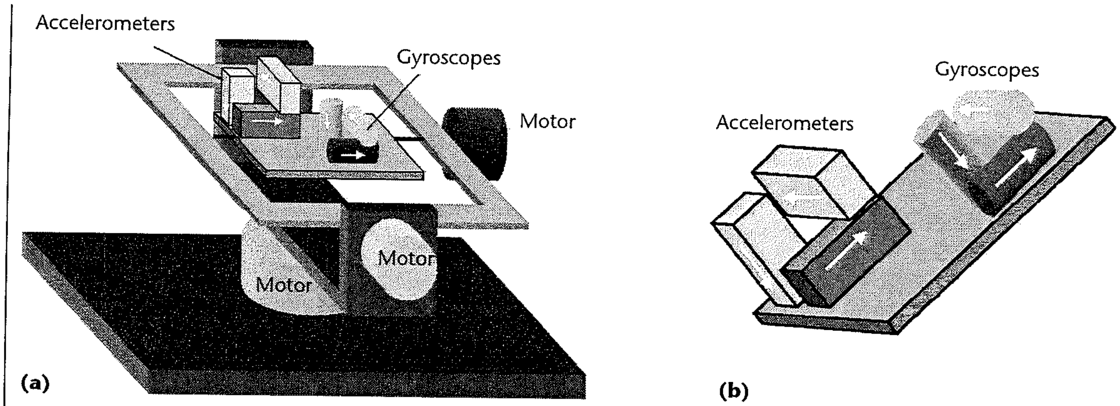
- *Complete*—tracking all six degrees of freedom (position and orientation);
- *Accurate*—with resolution better than 1 mm in position and 0.1 degree in orientation;
- *Fast*—running at 1,000 Hz with latency less than 1 ms, no matter how many ToCs are deployed;
- *Immune to occlusions*—needing no clear line of sight to anything else;
- *Robust*—resisting performance degradation from light, sound, heat, magnetic fields, radio waves, and other ToCs in the environment;
- *Tenacious*—tracking its target no matter how far or fast it goes;
- *Wireless*—running without wires for three years on a coin-size battery; and
- *Cheap*—costing $1 each in quantity.

If this magic ToC existed, we would use it for everything. The reality is that every tracker today falls short on at least seven of these 10 characteristics, and that number is unlikely to shrink much in the foreseeable future.

But all is not lost! Researchers and developers have pragmatically and cleverly exploited every available physical principle to achieve impressive results for specific applications. We'll start with an overview of some of the available ammunition and the strengths and weaknesses of each and then look at some specific applications and the tracking technologies that have been employed successfully in each.

Available ammunition

Although designers have many pose estimation algorithms to choose among, they have relatively few sensing technologies at their disposal. In general, the technologies sense and interpret electromagnetic fields or waves, acoustic waves, or physical forces. Specifically, motion tracking systems most often derive pose estimates from electrical measurements of mechanical,



1 **(a) Stable-platform (gimbaled) INS. (b) Strapdown INS.**

inertial, acoustic, magnetic, optical, and radio frequency sensors.

Each approach has advantages and limitations. The limitations include modality-specific limitations related to the physical medium, measurement-specific limitations imposed by the devices and associated signal-processing electronics, and circumstantial limitations that arise in a specific application. For example, electromagnetic energy decreases with distance, analog-to-digital converters have limited resolution and accuracy, and body-worn components must be as small and lightweight as possible. Although alternative classifications are possible, we discuss the available ammunition using a traditional medium-based classification.

Mechanical sensing

Arguably the simplest approach conceptually, mechanical sensing typically involves some form of a direct physical linkage between the target and the environment. The typical approach involves an articulated series of two or more rigid mechanical pieces interconnected with electromechanical transducers such as potentiometers or shaft encoders. As the target moves, the articulated series changes shape and the transducers move accordingly. Using a priori knowledge about the rigid mechanical pieces and online measurements of the transducers, you can estimate the target's position (one end of the link) with respect to the environment (the opposite end).

This approach can provide very precise and accurate pose estimates for a single target, but only over a relatively small range of motion—typically one cubic meter. In his pioneering HMD work in 1968, Sutherland built a mechanical tracker composed of a telescoping section with a universal joint at either end. While Sutherland and his colleagues found the system too cumbersome in practice, they relied on it as a "sure method" of determining head pose. The most common uses of mechanical sensing today are for boom-type tracked displays that use counterweights to balance the load and for precision 3D digitization over a small area. Commercial examples include the Boom 3C by FakeSpace and the FaroArm by Faro Technologies.

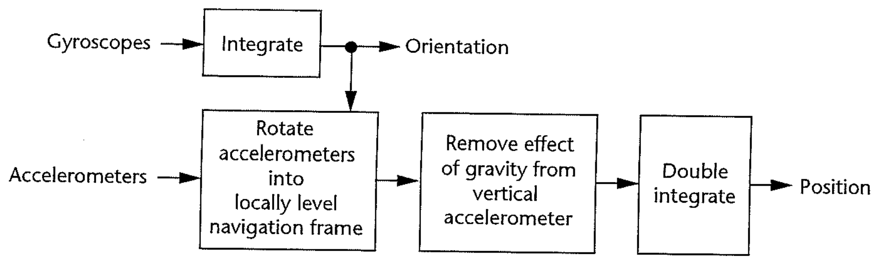
Articulated haptic devices such as the Phantom by SensAble Technologies inherently include mechanical tracking of the force-feedback tip. These devices need to know the tip position to apply appropriate forces, and the electromechanical devices typically used to provide the forces can also be used to sense the position.

Inertial sensing

Inertial navigation systems (INSs) became widespread for ships, submarines, and airplanes in the 1950s, before virtual reality or computer graphics were even conceived, but they were the last of the six ammunition technologies to be introduced for computer graphics input devices. The reason is straightforward: an INS contains gyroscopes, and early high-accuracy spinning-wheel gyroscopes weighed far too much to be attached to a person's body. Not until the advent of MEMS (microelectronic mechanical systems) inertial sensors in the 1990s did the development of inertial input devices begin.

Originally, inertial navigation systems were built with a gimbaled platform (see Figure 1a) stabilized to a particular navigation reference frame (such as north-east-down) by using gyroscopes on the platform to drive the gimbal motors in a feedback loop. The platform-mounted accelerometers could then be individually double-integrated to obtain position updating in each direction, after compensating for the effect of gravity on the vertical accelerometer. Most recent systems are of a different type, called strapdown INS (see Figure 1b), which eliminates mechanical gimbals and measures a craft's orientation by integrating three orthogonal angular-rate gyroscopes strapped down to the craft's frame. To get position, three linear accelerometers, also affixed to the moving body, measure the acceleration vector in body-frame, which is then rotated into navigation coordinates using the current rotation matrix as determined by the gyroscopes. The result is a navigation-frame acceleration triad just like that measured by the accelerometers in the stable-platform INS, which can be gravity-compensated and double-integrated in the same way. Figure 2 illustrates this flow of information.

Inertial trackers might appear to be the closest thing to a silver bullet of all the ammunition technologies we describe here. Gyroscopes and accelerometers are already available in chip form, and within the next decade we expect to see a single-chip six-axis strapdown inertial measurement unit—that is, with three gyroscopes and three accelerometers. Inertial sensors are completely self-contained, so they have no line-of-sight



2 Basic strap-down inertial navigation algorithm.

Tracking Performance Specifications and Requirements

In deciding the quality of tracking required for an application involving visual simulation such as virtual reality, there are several possible goals:

- The user feels presence in the virtual world.
- Fixed virtual objects appear stationary, even during head motion (perceptual stability).
- No simulator sickness occurs.
- Tracking artifacts don't affect task performance.
- Tracking artifacts remain below the detection threshold of a user looking for them.

Several types of tracking errors can contribute in varying degrees to destroying the sense of presence or perceptual stability, causing sickness, or degrading task performance. Various authors and manufacturers have focused on different specifications or defined them differently, and every type of tracker has its own complicated idiosyncrasies that would require a thick document to characterize in complete detail. However, Table A presents six specifications that can capture the essential aspects of tracking performance that affect human perception of a virtual environment while a tracked object is still (static) or moving (dynamic).

There's no clearly defined distinction between spatial jitter and creep, as they could be thought of representing the high- and low-frequency portions of a continuous noise spectrum. A reasonable cutoff might be to designate as creep any motion slower than a minute hand in orientation (0.1 degree per second) and slower than 1 mm per second in translation, with everything else called jitter.

Table A. Tracking performance specifications.

Static	Dynamic
Spatial distortion. Repeatable errors at different poses in the working volume, including effects of all sensor scale factors, misalignments, and nonlinearity calibration residuals, and repeatable environmental distortions.	*Latency.* The mean time delay after a motion until corresponding data is transmitted. It's possible to specify the latency of the tracker and other subsystems separately, but they don't simply add up.
Spatial jitter. Noise in the tracker output that causes the perception of the image shaking when the tracker is actually still.	*Latency jitter.* Any cycle-to-cycle variations in the latency. When moving, this will cause stepping, twitching, multiple image formation, or spatial jitter along the direction the image is moving.
Stability or creep. Slow but steady changes in tracker output may appear over time. The cause might be temperature drift or repeatability errors if the tracker is power-cycled or moved and returned to the same pose.	*Dynamic error (other than latency).* This error type includes any inaccuracies that occur during tracker motion that can't be accounted for by latency or static inaccuracy (creep and spatial distortion). This might include overshoots generated by prediction algorithms or any additional sensor error sources that are excited by motion.

requirements, no emitters to install, and no sensitivity to interfering electromagnetic fields or ambient noise. They also have very low latency (typically a couple of milliseconds or less), can be measured at relatively high rates (thousands of samples per second), and measured velocity and acceleration can generally be used to predict the pose of a head or a hand 40 or 50 ms into the future. Good inertial sensors also offer extremely low jitter (see the sidebar, "Tracking Performance Specifications and Requirements").

The weakness that prevents inertial trackers from being a silver bullet is *drift*. If one of the accelerometers has a bias error of just 1 milli-g, the reported position output would diverge from the true position with an acceleration of 0.0098 m/s^2. After a mere 30 seconds, the estimates would have drifted by 4.5 meters! If you

look closely at Figure 2, you can see that an orientation error of 1 milliradian coming from the gyroscopes would produce a gravity compensation error of 1 milli-g on one of the horizontal accelerometers, causing just this calamity.

Even very good gyroscopes (the kind you wouldn't want to wear on your head) drift by a milliradian within a short time. Nevertheless, given the advantages we've enumerated, inertial sensors can prove very valuable when combined with one or more other sensing technologies, such as those we describe next. Inertial sensors have provided the basis for several successful hybrid systems.

Acoustic sensing

Acoustic systems use the transmission and sensing of sound waves. All known commercial acoustic ranging systems operate by timing the flight duration of a brief ultrasonic pulse.

In contrast, in 1968 Sutherland built a continuous carrier-phase acoustic tracking system to supplement his mechanical system.[7] This system used a continuous-wave source and determined range by measuring the phase shift between the transmitted signal and the signal detected at a microphone. Meyer and colleagues point out that this "phase-coherent" method enables continuous measurement without latency but can only measure relative distance changes within a cycle.[3] To measure absolute distance, you need to know the starting distance and then keep track of the number of accumulated cycles. Another problem, which could be the reason no successful implementation of the phase-coherent approach has been developed, is the effect of multipath reflections. *Multipath*, a term also associated with radio transmission, indicates that the signal received is often the sum of the direct path signal and one or more reflected signals of longer path lengths. Because walls and objects in a room are extremely reflective of acoustic signals, the amplitude and phase of the signal received from a continuous-wave acoustic emitter in a room will vary drastically and unpredictably with changes in the receiver's position.

An outstanding feature of pulsed time-of-flight acoustic systems is that you can overcome most multipath reflection problems by waiting until the first pulse arrives, which is guaranteed to have arrived via the direct path unless the signal is blocked. The reason this method works for acoustic systems but not for radio frequency and optical systems is that sound travels relatively slowly, allowing a significant time difference between the arrival of the direct path pulse and the first reflection.

Point-to-point ranging for unconstrained 3D tracking applications requires transducers that are as omnidirectional as possible, so that the signal can be detected no matter how the emitter is positioned or oriented in the tracking volume. To achieve a wide beam width, you must use small speakers and microphones with active surfaces a few millimeters in diameter. This is convenient for integration into human motion tracking devices and helps reduce off-axis ranging errors, but the efficiency of an acoustic transducer is proportional to the active surface area, so these small devices can't offer as much range as larger ones.

To improve the range, most systems use highly resonant transducers and drive them with a train of electrical cycles right at the resonant frequency to achieve high amplitude. This results in a received waveform that "rings up" gradually for about 10 cycles to a peak amplitude then gradually rings down. For a typical envelope-peak detection circuit, this means the point of detection is delayed about 10 cycles—about 90 mm—from the beginning of the waveform. By detecting on the second or third cycle instead of the 10th, you can greatly reduce the risk of multipath reflection.

In our experience, this is one of the most important issues for accurate ultrasonic tracking outside of controlled laboratory settings, and it is the crux of how InterSense's ultrasonic ranging technology remains accurate at longer ranges than others.

The physics of ultrasonic waves in air and transducer design dictate other design trade-offs and considerations as well. Most ambient noise sources fall off rapidly with increasing frequency, so operating at a higher frequency is beneficial for avoiding interference, and the shorter wavelengths offer higher resolution. However, selecting a higher frequency reduces the range because of problems with transducer size and frequency-dependent attenuation of sound in air, which starts to play a significant role by 40 kHz and becomes the dominant factor in limiting range by 80 kHz, depending on humidity.

Ultrasonic trackers typically offer a larger range than mechanical trackers, but they're not a silver bullet. Their accuracy can be affected by wind (in outdoor environments) and uncertainty in the speed of sound, which depends significantly on temperature, humidity, and air currents. A rule of thumb is that the speed of sound changes about 0.1 percent per degree Fahrenheit of temperature differential. This corresponds to about a one-millimeter error per degree Fahrenheit at one meter.

Acoustic systems' update rate is limited by reverberation. Depending on room acoustics and tracking volume, it may be necessary for the system to wait anywhere from 5 to 100 ms to allow echoes from the previous measurement to die out before initiating a new one, resulting in update rates as slow as 10 Hz. The latency to complete a given acoustic position measurement is the time for the sound to travel from the emitter to the receivers, or about one millisecond per foot of range. This is unaffected by room reverberation and is usually well under 15 ms in the worst case. However, in a purely acoustic system with a slow update rate, the need to wait for the next measurement also affects system latency.

Acoustic systems require a line of sight between the emitters and the receivers, but they're somewhat more tolerant of occlusions than optical trackers (which we discuss later) because sound can find its way through and around obstacles more easily. Finally, we have yet to see a purely acoustic tracker that doesn't go berserk when you jingle your keys.

You can address most of the shortcomings we've mentioned by building a hybrid system that combines acoustic sensors with others that have complementary characteristics—inertial sensors, for example.

Magnetic sensing

Magnetic systems[8] rely on measurements of the local magnetic field vector at the sensor, using magnetometers (for quasi-static direct current fields) or current induced in an electromagnetic coil when a changing magnetic field passes through the coil (for active-source alternating current systems). Three orthogonally oriented magnetic sensors in a single sensor unit can provide a 3D vector indicating the unit's orientation with respect to the excitation.

You can use the earth's magnetic field as a naturally occurring, widely available DC source to estimate heading. The shape of the earth's magnetic field varies to some extent over the planet's surface, but you can use a look-up table to correct for local field anomalies.

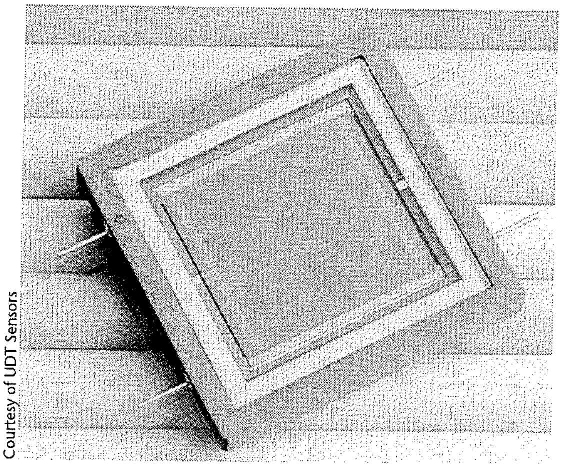
Alternatively, you can actively induce excitations with a multicoil source unit. This has been a popular means for tracking for interactive graphics for many years. You can energize each of the source unit coils in sequence and measure the corresponding magnetic field vector in the sensor unit. With three such excitations, you can estimate the position and orientation of the sensor unit with respect to the source unit.

However, ferromagnetic and conductive material in the environment can affect a magnetic field's shape. A significant component of the resulting field distortion results from unintended fields that appear around nearby conductive objects as the source induces eddy currents in them. These small fields act in effect as small unwanted source units. The most common approach to addressing these distortions is to ensure that the working volume contains no offending objects. This is why, for example, you might see a projector-based display system built out of wood or plastic. If you can't eliminate the offending objects (perhaps because they're an integral part of the application) you can try to model and correct for the resulting distortions.

You can use alternating or direct current signals to excite the source unit's coils. The use of AC was initially popular, but precisely because of the transient distortions we just mentioned, manufacturers introduced the use of DC fields. Even with DC fields, you must wait for the initial transient of each excitation to subside. Furthermore, you must make an additional excitation-free measurement of the ambient magnetic field to remove its effect.

With both AC and DC active source systems, the useful range of operation is severely limited by the inverse cubic falloff of the magnetic fields as a function of distance from the source. Position resolution in the radial direction from source to sensor depends on the gradient of the magnetic field strength, and thus the positional jitter grows as the fourth power of the separation distance.

Despite magnetic field strength and distortion problems, there are three noteworthy advantages to a magnetic approach to tracking humans. First, the size of the user-worn component can be quite small. Second, magnetic fields pass right through the human body, eliminating line-of-sight requirements. Third, you can use a single source unit to simultaneously excite (and thus track) multiple sensor units.


Courtesy of UDT Sensors

3 **Position sensing detector.**

Optical sensing

Optical systems rely on measurements of reflected or emitted light. These systems inevitably have two components: light sources and optical sensors. The light sources might be passive objects that reflect ambient light or active devices that emit internally generated light. Examples of passive light sources include distinguishable colored fiducials and even the natural surfaces in the environment. Examples of active light sources include light-emitting diodes (LEDs), lasers, or simple light bulbs.

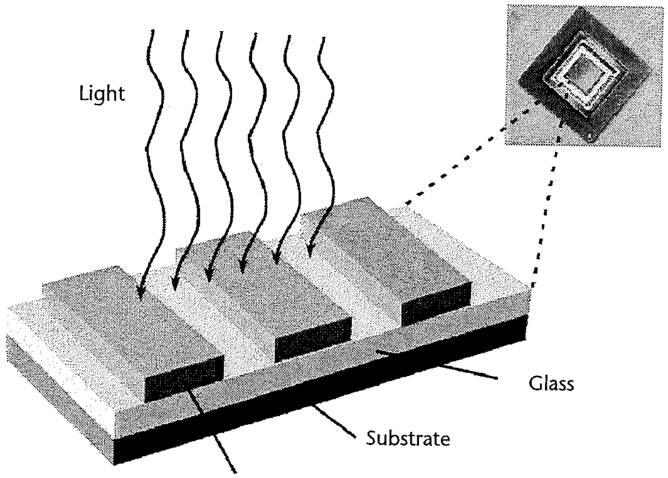
Optical sensors can be either analog or digital devices. Analog sensors offer continuous voltages indicating the overall intensity or centroid position of the aggregate light reaching the sensor. Digital sensors offer a discrete image of the scene projected onto the sensor. Both types of devices can be 1D or 2D. One-dimensional sensors can typically be sampled and processed at a higher rate than 2D sensors, but 2D sensors offer more information per (complete) sample. (Later, we'll describe some systems that use 1D optical sensors and some that use 2D sensors.)

Lenses and apertures can be used to project images onto the sensor, indicating the angle to the source. You can also use the intensity of light reaching the sensor to estimate the distance to the source. Filters can be added to selectively admit or reject certain wavelengths of light. For example, a sensor system might use infrared light sources in conjunction with filters that only admit infrared light, effectively providing a light "channel" separate from the ambient visible light.

The simplest analog sensor is a photosensor, a device that simply changes resistance as a function of the quantity of light reaching it. While individual photosensors offer relatively little information, relative or ratiometric amplitudes within a set of sensors can offer position information. Photosensors have the advantage of simplicity and speed.

An analog position sensing detector (PSD) is a 1D or 2D semiconductor device that produces a set of currents that indicate the position of the centroid of the light reaching the sensor (see the example in Figure 3). Like photosensors, PSDs offer measurements based on the total light reaching the device. As such, the target light source amplitude is typically under program control, so

FOXLIN0003010



4 Simplified diagram of some cells from an image-forming charge-coupled device (CCD).

Outside-In or Inside-Out?

When using optical emitters and sensors for tracking, we must consider whether to put the light sources on the moving target and the sensors in the environment, or vice versa. The first of these two alternatives is often described as *outside-looking-in,* the second as *inside-looking-out.*

However, these terms can be misleading. For example, the Vulcan Measurement System from Arc Second (http://www.arcsecond.com) employs multiple optical sensors mounted on the target and two or more spinning light sources mounted in the environment. The spinning light sources sweep out distinct planes of light that periodically hit the optical sensors, and the system uses the timing of the hits to derive the sensors' positions. While the target-mounted optical sensors do indeed "look outward" toward the environment, the system actually has the orientation sensitivity characteristics of what is typically called an outside-looking-in system. Thus, the typical inside-looking-out characterization would be misleading.

The actual distinguishing factor is whether bearing angles to reference points are measured from the outside or the inside.

that the system can use differential signaling to distinguish the target from the ambient light.

The more familiar digital image-forming devices such as charge-coupled devices (CCDs) typically use a dense 1D or 2D array of pixel sensors that convert light energy (photons) into an electrical charge. These systems use the array of pixel sensors to produce a discretely sampled image of a scene by simultaneously opening the pixel sensors to collect light energy over a short time interval. Electronics surrounding the pixels then transfer the array of charges off the chip. Figure 4 is a simplified diagram of a CCD.

Although a large set of pixel sensors can be triggered simultaneously, measuring and transferring the per-pixel charge into a computer can be relatively time-consuming. The result is that image-forming devices are typically limited to relatively few measurements per unit of time when compared to the simpler analog optical PSD described earlier.

Of course, 1D or 2D images typically offer more constraints on a pose estimate—for example, letting you extract shape, shading, or motion of multiple image features. However, you must interpret the image to obtain those constraints, a process that can be computationally costly. Special-purpose processing can help, but interpretation is still difficult because of variations in lighting and surface properties, occlusions, and independent (confounding) object motion in the images.

As with other types of sensors, you can combine measurements from two or more optical sensor units to obtain more information than you could get from a single sensor unit. The most common example is the use of multiple sensor units in known locations to estimate the position of a light source with respect to the sensor units. Related methods include triangulation and multibaseline correlation.

When using multiple optical sensors in this fashion, you must consider whether to put the light sources on the moving target and the sensors in the environment, or vice versa. These two alternatives are often referred to as *outside-looking-in* and *inside-looking-out* respectively, although that characterization can be misleading (see the sidebar, "Outside-In or Inside-Out?"). Outside-looking-in has the advantage that the target-borne component is relatively simple—a small retroreflector, for example. For user tracking this is particularly attractive. The disadvantage is that if you want to know the target's orientation, you must estimate it from the relative measured positions of multiple target-borne light sources, and this relative measure is sensitive to the distance of the fixed sensors. Acoustic systems have the same problem, and in each case you must be mindful of the sensitivity of the measurements given the choice of source and sensor locations.

The primary disadvantage of all optical systems is that there must be a clear line of sight between the source and the sensor. For analog photosensors or PSDs, a partial occlusion may turn out to be the biggest problem, as it results in a signal that's plausible but incorrect. Image-forming devices can more readily recognize and reject partially occluded sources based on shape, but they must cope with potentially unknown features and with signal strengths and error models that are often difficult to predict.

Analog optical PSD sensors combined with active light sources offer the combination of relatively high spatial precision and update rates. For tracking applications in particular, this combination can be valuable. Passive systems with image-forming devices, on the other hand, offer the hope of operating in natural or unadorned environments. (Most environments you work in have some interesting visible features.) Furthermore, in cases such as real-time computer-aided surgical systems, where

graphical registration is critical, image-forming devices can provide closed-loop feedback between the real environment and the tracking system.

Radio and microwave sensing

Radio and microwaves haven't been exploited much in tracking human motion, but they're widely used in navigation systems and various airport landing aids and radar systems. These technologies have also begun to find application in local positioning systems that find radio frequency asset tags in warehouses or hospitals—they might well be used for human motion tracking systems in the future as the precision improves and the technology becomes smaller and cheaper.

Electromagnetic wave-based tracking techniques can provide vastly greater range than quasi-static magnetic fields because radiated energy in a field of radius r dissipates as $1/r^2$, whereas the dipole field strength gradient drops off as $1/r^4$. Furthermore, radio waves suffer negligible absorption losses in air and are virtually unaffected by wind and air temperature, so they're uncompromised outdoors or in large open spaces where acoustic systems have difficulty. Unfortunately, radio waves are rapidly attenuated in water, so the human body is opaque to all radio frequencies useful for precision ranging.

Most radio navigation systems operate on the principle of time-of-flight range finding, much like the acoustic systems we described earlier. The waves travel about a million times faster (roughly 1 foot/ns, as opposed to 1 foot/ms for sound), making the task of measuring time of flight with sufficient precision much more difficult. For example, ranging with 1 mm resolution would require a timer that can count at 300 GHz, implying expensive and power-consuming electronics.

However, various signal-processing strategies can provide high resolution without brute-force fast counting. For example, the Global Positioning System (GPS) uses a delay-locked loop (DLL) to keep adjusting the delay τ to maximize the correlation of the incoming signal with a locally generated replica of it. Unfortunately, this and any other scheme that uses sinusoidal signals with long wavelengths is vulnerable to multipath distortions in indoor applications, where reflections are abundant and unavoidable. The only radio frequency-based precision motion tracker we know of achieved a surprisingly good resolution of a few millimeters, but it used large racks of microwave equipment and was typically demonstrated in an empty room covered on five sides with thick, radio frequency-damping foam blocks.

Considerable interest has risen recently in ultra-wideband (UWB) ranging, which uses nonsinusoidal electromagnetic signals such as impulses. UWB can't be allocated specific regions of the radio spectrum as are conventional radio frequency systems, because it necessarily emits energy across the whole spectrum from DC to several GHz. However, because the energy is spread out over this wide spectrum, UWB emissions tend to look like very low-level background noise. This generally reduces the risk of UWB systems interfering with other (narrowband) systems, and the US Federal Communications Commission recently approved the use of UWB within certain power constraints. The outstanding advantage of the UWB paradigm is the improved ability to reject multipath signals. With pulses as short as 200 picoseconds, all reflection paths delayed by 6 cm or more can be easily disregarded. For this reason, it seems that UWB is more likely than other radio frequency technologies to find applications in indoor ranging, but it's still too early to say.

An arsenal of application-specific approaches

Alas, among our available ammunition we find no silver bullet—no single modality, technology, or device that can overcome the problems that arise for any tracking application. However, as Brooks says in his original article, "There is no royal road, but there is a road."[6] The key is to recognize the different needs of specific applications and match those needs with one or more of the available technologies. For example, the approaches you would use for tracking a person's limbs for a computer animation would differ from those you would use for tracking a surgeon's head in a surgical suite. Researchers and engineers have had significant success developing tracking systems for specific applications or classes of applications. Here we look at some of the successes related to computer graphics.

Head tracking for interactive graphics

Head tracking for interactive computer graphics presents arguably the biggest challenge in motion tracking. We humans have a lifetime of experience in perceiving our environment and interacting with physical objects. As a result, fooling the human senses can prove exceedingly challenging, requiring high spatial accuracy and resolution, low latency, and high update rates. In the lab, where significant infrastructure is possible, researchers and engineers have succeeded in developing both general-purpose, small-scale tracking systems and relatively high-performance, wide-area systems such as the 3rdTech HiBall-3000 and the InterSense IS-900, both of which we describe later.

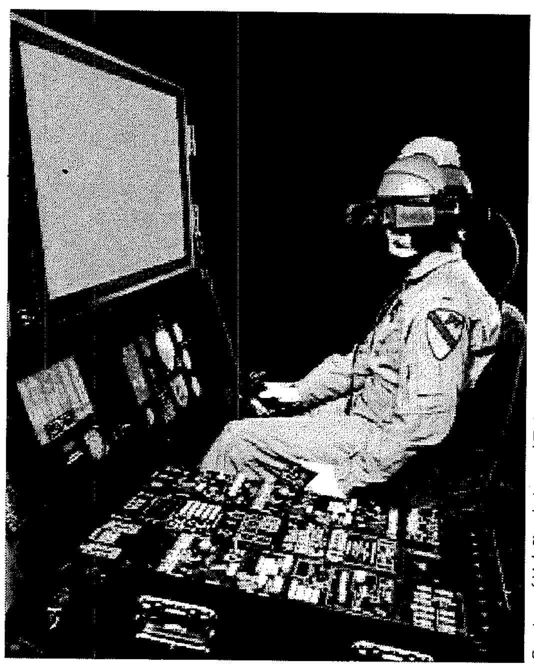
VR games, fly-throughs, and vehicle simulators

This is the class of applications that consumers typically think of when they think of virtual reality: a person seated at a desk or standing in a pod at an arcade wears a fully immersive HMD and looks around inside a virtual world (see Figure 5, next page). The graphics presented in the HMD are updated in response to head movements using a commercial magnetic head tracker. Because players can't actually walk around the virtual world, they virtually "fly" through it. If seated, the player can use a joystick to control flying direction and speed; if standing, the player does this more conveniently using a second free-air tracking device in the hand, or by simply looking in the desired direction of travel and holding down a button to fly forward.

For entertainment-based flight or driving simulators, this fly-through or drive-through paradigm works out naturally; for other applications it's less convincing than a real walk-through but much cheaper, especially in terms of tracking equipment and space requirements.

FOXLIN0003012

5 Virtuality 2000SU location-based entertainment pod. The head-mounted display and hand controller both have magnetic sensors that are tracked by a magnetic source unit concealed in the rim, which also keeps the user from straying too far from the source.



6 AVCATT-A simulator.

Courtesy of Link Simulation and Training

These applications are the easiest from the tracking perspective, as they require neither a large tracking area, extremely high accuracy, nor the tracking of large numbers of objects. They have historically been handled using magnetic trackers, and with good reason: the jitter and interference problems with magnetic trackers that we described earlier are extremely range dependent. At very close range, magnetic trackers can perform quite well, and they're moderately priced and easier to use than most of the more elaborate trackers we describe later. The absence of line-of-sight requirements is extremely convenient and unique to magnetic trackers.

Polhemus trackers have been around since the dawn of VR. IsoTrak was used in the first commercial VR systems from VPL, W Industries/Virtuality Systems, and Division (all now defunct). Since then, both Polhemus and Ascension Technology have put forward second-generation magnetic trackers that cut the original latency by an order of magnitude.

In applications in which the player doesn't move around much in position, a three-degrees-of-freedom (DOF) orientation tracker often suffices. This still lets a player look around freely in the virtual world and fly in the direction of gaze by pressing a button (or to use a joystick). The main thing lost compared to a six-DOF system is the subtle motion parallax effect when the player moves her head a few inches left or right.

However, if you can get away with three-DOF, orientation-only tracking, there are trackers much less expensive and even simpler to use than magnetic trackers. These are called "sourceless" trackers—all you have to do is attach a sensor to the HMD and plug it into the computer.

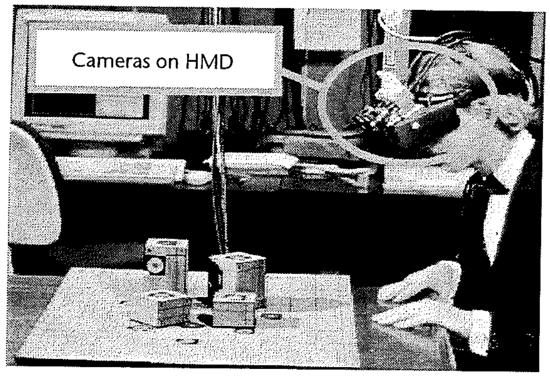
Early sourceless trackers consisted of just an inclinometer to measure pitch and roll and a compass for yaw. These suffered from a terrible degree of slosh as the fluid in the inclinometer tried to reestablish level after the player's head moved. Switching to solid-state, accelerometer-type inclinometers did nothing to help this problem, because all inclinometers work based on gravity and therefore must be sensitive to even small horizontal head accelerations as well. InterSense's Gyroscopic Earth-Stabilized Orientation Sensing (GEOS) technology overcomes this by using microgyroscopic sensors to measure all three DOFs and automatically correcting any drift produced by the gyroscopes using an inclinometer and compass when it detects that the person's head is still. If consumer HMDs ever take off, this will likely be the tracking technology of choice, because it can be built into the headset, requires no user setup, and provides high-resolution, low-latency, low-cost tracking with no environmental sensitivities.

For high-end military flight simulators, magnetic trackers have too much latency and distortion arising from metals in the environment. As such, these applications often use optical or inertial approaches. For example, the AVCATT-A simulator (Figure 6) uses the InterSense SimTracker, an acoustic-inertial hybrid that has a prediction option to address rendering latency.

VR walk-throughs

When a VR system's goal is to simulate the experience of walking or running around in a virtual world, as in architectural walk-throughs or dismounted infantry simulations (DIS), nothing can beat an actual walk-through VR system for creating the illusion of presence. This requires a wide-area tracking capability (room size or even gymnasium size)—a tremendous technical challenge that has fired the imaginations of a few ambitious tracker designers, including us. The difficulties of building a scalable-area precision tracking system are so large, and the market as yet so undeveloped, that we're only aware of two commercially available solutions: the University of North Carolina optical ceiling tracker,[9]

7 **UNC-Chapel Hill is exploring the use of augmented reality to assist with breast biopsy procedures.**



8 **Video-based head-mounted display for augmented reality.**

brought to market as 3rdTech HiBall-3000, and the Massachusetts Institute of Technology acoustic-inertial Constellation,[10] brought to market as InterSense IS-900. They both achieve scalable range by using a ceiling-mounted array of reference emitters, and in both systems, the emitters are mounted in strips that clip to ceiling grids for easier installation. The HiBall is a purely electro-optical system, and it achieves its low latency by making measurements at a rapid 2,000-Hz rate. In contrast, the IS-900 uses a less-dense array of lower-update-rate acoustic emitters, but maintains its low latency, robustness, and predictive capability by fusing these range measurements with inertial sensors.

No publications have yet compared the performance of the two systems, but we expect the HiBall to have the upper hand in terms of accuracy, while the IS-900 has a somewhat less dense infrastructure, lighter and smaller head-mounted sensors, and a wireless option.[11] However, we haven't found much interest in wireless tracking for use with HMDs because there are no high-resolution HMDs with wireless capability yet. In fact, the difficulty of making even one high-resolution stereoscopic wireless video link operate reliably over a large space (let alone supporting multiple users in the same space) might be so great that by the time walk-through VR systems emerge as a real deployable application, they'll be based on wearable computers that generate the graphics on-body to avoid the need for transmitting video. For this type of system, neither the HiBall nor the IS-900 is really appropriate, and a wearable self-tracking system with no active infrastructure, as we discuss later, would be much more easily integrated.

Augmented reality

In VR, a person's view of the real world is replaced completely by computer graphics. The basic idea behind augmented reality[11-13] is to use special HMDs to add 3D computer graphics to a person's view of the real world, so that the virtual objects appear to coexist with physical objects. One application of AR is to provide medical specialists with what's effectively "X-ray vision." For example, a physician could use a tracked head-worn display to view live ultrasound or laparoscope data in 3D, accurately registered on the appropriate part of the patient's body. Fuchs and colleagues at UNC-Chapel Hill are working on such systems to aid both breast biopsy and laparoscopic procedures (see Figure 7).

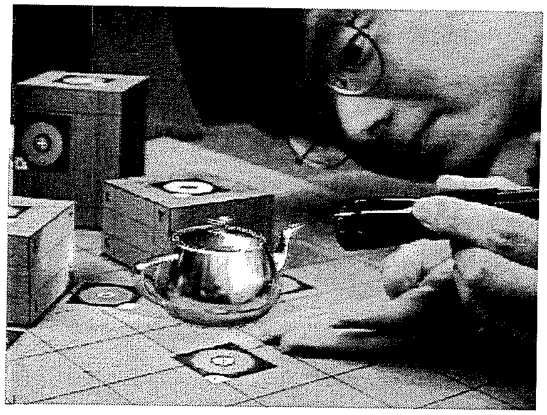
Another application is AR outdoors. The idea is to develop handheld or head-worn, pose-aware AR devices for soldiers, tourists, and others; these devices would provide visual information and icons visually registered with objects in the real world.[13]

Beyond graphics, projects such as Nexus[14] are aimed at global location-aware networks that associate information with specific geographical locations. Someday your personal digital assistant might function as a pose-aware magic lens, overlaying useful information on the physical world around you. You might select "Bus Stops" from a menu and then, pointing the PDA down the street, you would see the live scene augmented with static icons fixed on the bus stops (which you would click to get a schedule) and moving labels attached to approaching buses.

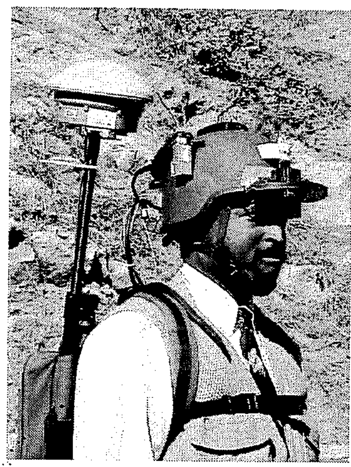
Tracking your HMD or a PDA with the position and orientation accuracy necessary to geospatially register graphics with the real world is difficult. Doing this outdoors is extremely difficult. Compared with immersive VR, AR applications are typically sensitive to static and dynamic tracking errors, not to mention a host of other calibration problems. Virtual objects that appear in the wrong location or jitter or "swim" around are immediately obvious when compared to static physical objects. This could clearly be serious for medical and military applications, but it would also be annoying if you were trying to use your PDA to navigate to the nearest Starbucks.

A common approach is to use cameras mounted on the display (as in Figure 8) to provide vision-based feedback, effectively closing the loop between the tracking and the display. Seminal work by Bajura and Neumann at UNC-Chapel Hill in the early 1990s used light-emitting diodes placed on the physical objects.[15] By tracking the locations of these clearly identifiable landmarks in the camera imagery, they were able to visually register virtual labels with the real objects. Later, State and colleagues at UNC-Chapel Hill used coin-size multicolored rings as passive landmarks, as shown in Figure 9 (next page).

More recent work attempts to do the same using natural features in the environment (such as the edges of physical objects) or the horizon as landmarks. Because video sensor rates are relatively low and information gar-



9 **Landmark tracking in action. The teapot is virtual.**



Courtesy of HRL Laboratories

10 **HRL Laboratories prototype person-portable mobile augmented reality system with GPS, inertial, and magnetic sensors.**

nered from passive landmarks is relatively noisy (compared to active beacons), researchers have been exploring hybrid systems that leverage the complementary characteristics of vision and other sensor technologies.[16] For example, InterSense is working on a vision-inertial hybrid for use in building-wide AR and robotics applications. Researchers such as Neumann at the University of Southern California and Azuma at HRL Laboratories have been pursuing vision-inertial hybrids for AR outdoors. Figure 10 shows the HRL Laboratories prototype; Figure 11 shows example results.

Head-and-hand tracking for projection-based interactive visualization and design

In the mid-1990s, the VR industry took a sharp change of course from HMD-based approaches to projection-screen systems such as the Cave Automatic Virtual Environment, VisionDome, and PowerWall, which offer more sociable viewing environments with less sweaty hardware to don. The requirements for tracking in these environments are quite different than for HMDs. With HMDs, the main requirement is for fast and smooth orientation tracking to produce the illusion that the virtual world is stationary while users turn their heads. In a projection-based virtual environment, the displays are affixed to the world and not the user. As a result, the images are less sensitive to head rotation, and the head-tracker orientation can be heavily filtered and relatively slow to respond, but the hand-tracked input device should be responsive. There are several important tracking considerations for a projection-based VE:

- *Ergonomics*. In addition to increased field of view, one major attraction of projection systems is the elimination of the bulky, sweaty HMD apparatus. A head tracker must attach to a light pair of glasses rather than a relatively substantial HMD. Hand-tracking devices must be comfortable to hold or wear for prolonged sessions.
- *Range*. Unlike most HMD applications, a typical projection system requires a 3 m × 3 m tracking area, and many VR theaters are much larger.
- *Wireless*. With groups of people milling around inside a projection environment, cables dragging on the floor are a tripping hazard, and they can also pull off users' stereo glasses. In addition, users frequently pass devices around, which is easier without wires.
- *Performance*. Position tracking with low latency and decent accuracy is necessary, so that a virtual representation of a tool appears to remain fixed to a handheld tracked wand as someone waves it around.

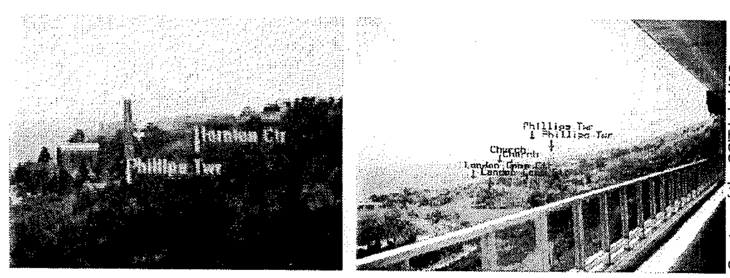
Figure 12 shows some of the tracked devices and the wireless links that have been developed for the IS-900 to adapt it for this new market.

Full-body motion capture

While VR and AR applications typically require precise position and orientation estimates for a single target—the user's head—applications such as biometrics and character animation require only position estimates, but for a large number of closely located targets—arms, elbows, hands, knees, feet, and perhaps a prop (a light saber or golf club, for example). Although the demands for absolute accuracy are arguably reduced, tracking many moving targets is a nontrivial task.

In the early 1970s, Woltring developed the Selspot

11 **Outdoor augmented reality results (low-resolution head-mounted display view) from HRL Laboratories and the Computer Graphics and Immersive Technologies Lab at USC.**



Courtesy of the CGIT Lab, USC

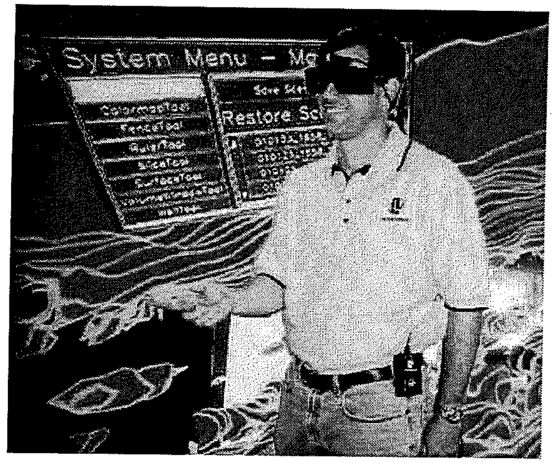
system for human motion studies. This system used several PSD analog optical sensors (see "Available Ammunition") fixed in the environment to track multiple LED sources on the user's body. This approach resembles the HiBall system mentioned earlier, but with a sensor–source role reversal. The Selspot system used online data collection and offline computation.

An interesting modern variation of the active LED approach is the Cartesian Opotoelectronic Dynamic Anthropometer system, or CODA, offered by Charnwood Dynamics. The system uses battery-powered, user-worn LEDs that are remotely triggered, and multiple 1D CCD digital image-forming camera sensor units fixed in the environment. The system is interesting in that it uses no lenses. Instead, it uses an optical grating with a pseudorandom bar pattern that effectively casts a distinctive bar shadow onto the sensor when a single LED is illuminated. Using high-speed digital signal processing techniques, the system finds the subpixel correlation of the grating pattern with the shadow and estimate the angle to the LED. The system uses multiple such 1D measurements to estimate the position of an LED, and multiple LEDs are excited in sequence to capture the user motion. The system can estimate six to 56 LED positions online, in real time, with relatively low latency (5 ms), at 800 targets per second for six LEDs and 100 targets per second for 56 LEDs.

The Ascension Technology ReActor is an active system with another interesting twist. The ReActor cleverly uses simple infrared detectors similar to the photosensors described earlier. This system uses multiple several-meter-long sensor units, each with an infrared detector embedded approximately every three centimeters, and infrared LEDs placed on the user's limbs. As the user moves around, the LEDs are sequentially energized, and each of the complete set of infrared detectors is sampled. The relative measured intensity along each sensor unit row indicates the position of the corresponding LED. This elegantly simple approach enjoys the speed afforded by the simple sensors, letting the system track many targets on multiple users. However, as with other systems using non-image-forming optical sensors, users must avoid light-colored or shiny surfaces in the operating area.

The ReActor's active user-worn targets can be made relatively unobtrusive. However, some companies make real-time motion capture systems with passive targets, typically small, lightweight, retroreflective markers. The markers are illuminated by infrared light synchronized with high-speed 2D digital imaging systems that can simultaneously image multiple targets. Of course, as we noted earlier, images must be processed to interpret the content, but the use of synchronized infrared light helps disambiguate the markers. Motion Analysis Corporation and Vicon are two companies offering such systems.

In general, optical systems can provide accurate estimates at relatively high speed and low latency with small user-worn components, but they all must deal with target occlusions—the disappearance and reappearance of targets in the sensor's view as the targets pass behind solid objects such as a body or a prop. Prior knowledge about the relationship between the moving targets can, by providing constraints on the motion, help the system survive momentary occlusions. For example, targets attached to your shoulder and arm will remain a fixed distance from each other. Adding more sensor units can help, too, by increasing overall visibility.

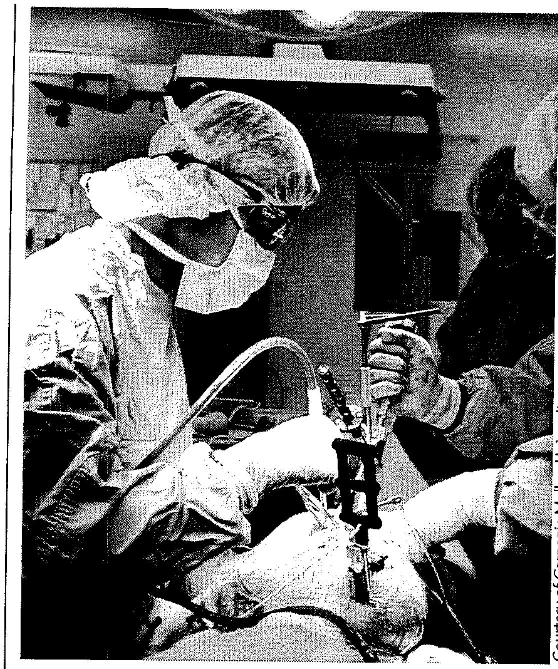


12 InterSense IS-900 wireless tracker option, including a belt-worn battery/radio module for the head tracker. For the wand, the battery and radio are built into the handle.

Some motion-capture systems use other ammunition besides light. Recall that magnetic fields pass right through your body. This makes a magnetic approach attractive because it eliminates the occlusion problems of optical systems. In addition, recall that you can track multiple magnetic sensor units by observing their simultaneous responses to excitations of a single source unit. These characteristics and the small size of the sensor units are all attractive. Both Polhemus and Ascension Technology offer wireless magnetic systems for motion capture. Measurand, the maker of ShapeTape, has a system that uses strips of special material that change electrical properties when they're bent. This technology lets you simultaneously monitor multiple joints for full motion capture.

Current research in this area is also interesting. While at the Naval Postgraduate School, Bachmann developed an inertial-magnetic hybrid system for motion capture. The approach, presented at the ACM Symposium on Virtual Reality Software and Technology in 2001, offers the advantage of being fast and sourceless.

At UNC-Chapel Hill, Vallidis and Bishop are currently working on a body-relative limb-tracking system that uses spread-spectrum audio.[17] The idea is to transmit pseudorandom audio signals from small speakers and then use digital signal processing to find the corresponding signal in the signals from small microphones. This correlation information indicates the delay and hence the distance between the source and sensor. Multiple such units can be used for body-relative limb tracking. The spread-spectrum approach offers immunity from environmental noise and a distribution of energy that minimizes the audible sound to a whisper. Although not sourceless, the approach offers an absolute measure of distances in a manner relatively immune to occlusions (low-frequency audio passes nicely around your body).


Courtesy of Carnegie Mellon University

13 Carnegie Mellon University's Robotics Institute and Shadyside Hospital in Pittsburgh use the HipNav system to assist in component placement during hip-replacement surgeries.

Tracking surgical instruments

Computer-aided surgery requires extremely precise tracking of surgical instruments, prosthetics, and even bones, in a relatively small controlled environment. Despite occlusion concerns, the ammunition of choice seems to be optical tracking. Popular commercial choices are the OptoTrak system by Northern Digital and the Flashpoint system by what was formerly Image Guided Technologies. For example, researchers at Carnegie Mellon University's Robotics Institute and Pittsburgh's Shadyside Hospital use OptoTrak for their HipNav system, which helps surgeons more precisely place replacement hip components (see Figure 13). Researchers at UNC-Chapel Hill use Flashpoint to track the ultrasound probe and biopsy needle for their image-guided breast biopsy AR system (Figure 7).

Camera tracking for media and entertainment

In the 1980s, Memorex introduced the catchy slogan "Is it live, or is it Memorex?" In a television commercial, a mother hears a piano and assumes it's her son practicing, when in fact it's a recording, and the boy is outside playing. Today, the question might be, "Is it real, or is it computer graphics?"

Tracking plays a critical role in today's television and cinema magic. For the position, size, and perspective of inserted computer graphics to match real objects in a scene, the camera pose (and optical parameters such as zoom) must be known throughout the entire sequence. And, because the problem is essentially AR, all the related problems noted earlier apply. Which approach producers choose depends chiefly on the amount of infrastructure they can afford. Typically, the use scenarios boil down to two situations: virtual sets where the computer graphics must be added live in real time or with little postprocessing—as in Fox News—and movie making where the computer graphics can be added later in postproduction—as in Sony Pictures' *Stuart Little 2.*

Besides temporal considerations, each situation has different infrastructure opportunities and constraints. With virtual sets, adding permanent tracking infrastructure is usually a reasonable option. Companies like Orad Hi-Tec Systems and InterSense have products that introduce or use infrastructure to track the camera. Orad, for example, offers two approaches. The first, InfraTrack, tracks the 6D pose of a camera using infrared LEDs on the camera and infrared sensors fixed in the studio. Like the landmark-based AR approaches described earlier, Orad's CyberSet product also does real-time vision-based tracking of known landmarks in the studio, using the television camera itself to extract patterns from a two-tone bluescreen background.

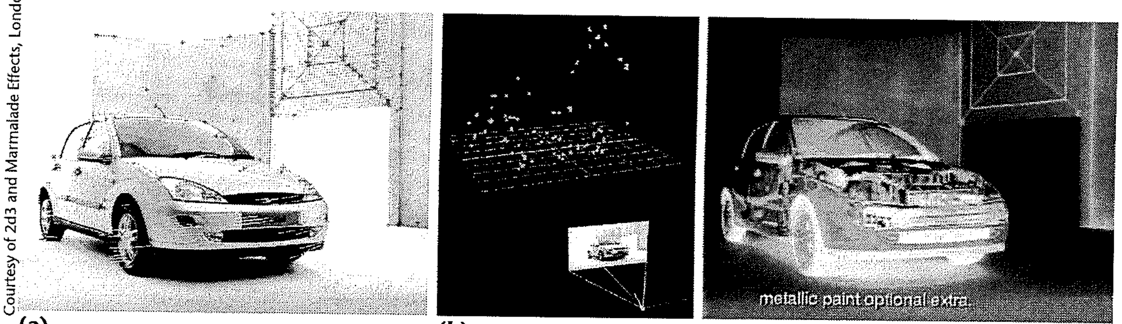
When making movies, it's typically impractical or impossible to add significant infrastructure to the environment, which might be an extemporaneous indoor or outdoor setting. Typically, technicians only have the scene's raw footage, usually collected with a moving camera, and some knowledge about the size and possibly the locations of some items appearing in the footage. To insert the computer graphics, they traditionally use a technique called match moving. The basic idea is to match the computer graphics camera view with the actual camera view. This involves estimating the 6D pose of the real camera with respect to visible 2D image features belonging to stationary objects in the real scene. Until recently, technicians would manually select suitable image features in the initial frames, run some form of image-based tracking over the remaining frames, and then compute a sequence of pose estimates minimizing some error function.

This sounds simple, but it isn't. For example, you don't know ahead of time which features belong to stationary objects, features come and go as they're occluded, false features appear at edges and edge junctions, lighting affects the appearance of the features, and on and on. Commercial products such as Boujou from 2d3 and MatchMover from RealViz now offer nearly automatic camera tracking. Boujou analyzes complete raw 2D footage, automatically finds and tracks (over time) features in the footage, and produces an optimal sequence of camera pose estimates, allowing technicians to add 3D computer graphics that match the 2D features in terms of perspective and motion. Figure 14 shows some examples provided by 2d3.

Offline match-moving or camera-tracking systems can typically operate noncausally—that is, they can look at an entire sequence of video, randomly scanning backward and forward in time to optimize the feature tracking. Online tracking systems don't have this luxury—they can only look at past measurements.

Cornelis and colleagues at Katholieke Universiteit Leuven in Belgium have been working on automatic estimation of camera motion, perspective camera calibration parameters, and complete 3D scene structure, using image sequences from a single, uncalibrated, freely moving camera.[18] Methods that attempt to simultaneously estimate 3D scene structure and camera motion face a difficult chicken-and-egg problem: to identify 2D image feature tracks that correspond to sta-

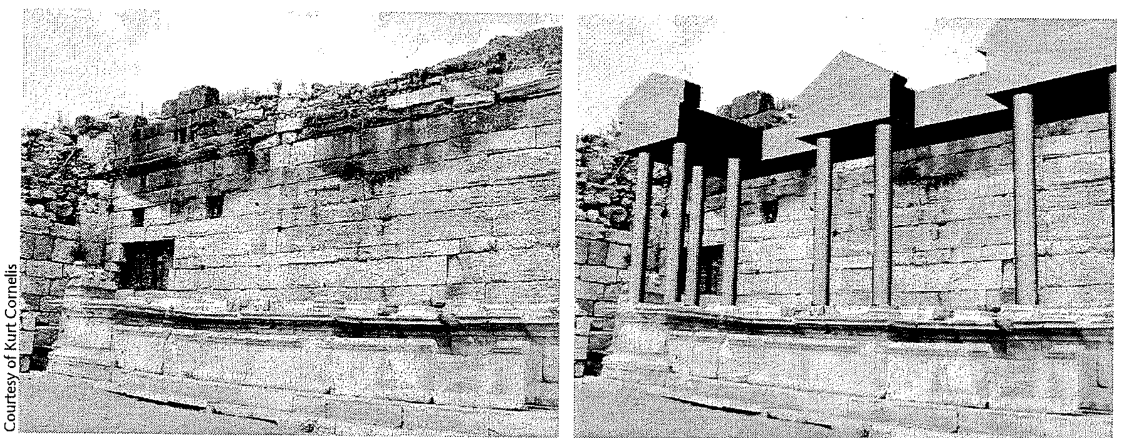
Courtesy of 2d3 and Marmalade Effects, London



(a) (b) (c)

14 **Boujou (2d3) examples from a Ford Focus commercial. (a) An annotated image from the original 2D footage. The red crosses indicate tracked features; the yellow lines indicate feature motion tracks. Note the different track directions throughout the scene—a result of camera rotation about the center of the car. (b) The 3D point cloud corresponding to the tracked features. (c) The final result combines actual 3D CAD models with the original 2D video to achieve a 3D X-ray effect.**

Courtesy of Kurt Cornelis



15 **Frame from a video presented by Kurt Cornelis and colleagues at the ACM Symposium on Virtual Reality Software and Technology, 2001. Fountain of the Upper Agora at Sagalassos in Turkey. Left: original footage. Right: final footage.**

tionary physical objects in the 3D scene, you need to know the camera motion; yet, to determine the camera motion you need to know which 2D image feature tracks correspond to stationary physical objects in the 3D scene. The approach of Cornelis and colleagues begins with a robust feature identification and tracking stage and then uses an iterative structure- and motion-recovery algorithm that attempts to solve for the 3D scene points that gave rise to the 2D feature tracks, the perspective camera model parameters, and the camera pose over time. Figure 15 shows some results of their approach for a video documentary on Sagalassos in Turkey. These researchers have structured their approach to run eventually online, in real time. The idea would be to give directors some means to assess the suitability and quality of raw footage as it's being collected, rather than waiting until postprocessing to discover problems.

Perhaps there will eventually be some convergence of these relatively unconstrained approaches and the more instrumented approaches described earlier for outdoor AR. This might offer a nice, shiny, stainless-steel bullet.

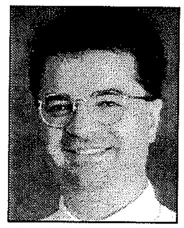
Conclusions

As savvy technologists, we sometimes arrive at preliminary solutions to a technological problem with great confidence and then later find out the problem is more complex than we thought. Motion tracking is no different from other problems in this respect, as both the essence and the accidents[6] make the problem quite challenging.

We hope our no-silver-bullet approach to the topic has resonated with you and that we've dispelled some of the myths surrounding the apparent black art of tracking, enlightening you to the admirable work that has been done and perhaps inspiring you to consider the range of possibilities for your application. Should you wish to delve more deeply into the subject, we particularly recommend, of course, our own recent surveys.[1,2] Although the two are similar in scope and worldview, the first provides much more tutorial coverage of the mathematical techniques involved in designing tracking systems,[1] whereas the latter attempts to comprehensively catalog the possible physical principles, what we've called the ammunition.[2] ■

References

1. B.D. Allen, G. Bishop, and G. Welch, "Tracking: Beyond 15 Minutes of Thought: SIGGRAPH 2001 Course 11," *Course Notes, Ann. Conf. Computer Graphics and Interactive Techniques* (Siggraph 2001), ACM Press, New York, 2001.
2. E. Foxlin, "Motion Tracking Technologies and Requirements," *Handbook of Virtual Environment Technologies*, chapter 8, K. Stanney, ed., Lawrence Erlbaum Publishers, Hillsdale, N.J., 2002, pp. 163-210.
3. K. Meyer, H. Applewhite, and F. Biocca, "A Survey of Position Trackers," *Presence: Teleoperators and Virtual Environments,* vol. 1, no. 2, 1992, pp. 173-200.
4. A. Mulder, *Human Movement Tracking Technology,* tech. report TR 94-1, School of Kinesiology, Simon Fraser University, Burnaby, B.C., Canada, 1994.
5. J. Richards, "The Measurement of Human Motion: A Comparison of Commercially Available Systems," *Proc. Fifth Int'l Symp. 3D Analysis of Human Movement*, Univ. of Tennessee, Chattanooga, Tenn., 1998; http://www.utc.edu/Human-Movement/3-d/proceds.htm.
6. F.P. Brooks, Jr., "No Silver Bullet—Essence and Accidents of Software Engineering," *Computer*, vol. 20, no. 4, Apr. 1987, pp. 10-19.
7. I.E. Sutherland, "A Head-Mounted Three Dimensional Display," *Proc. 1968 Fall Joint Computer Conf.,* vol. 33, part 1, Thompson Books, Washington, D.C., 1968, pp. 757-764.
8. F.H. Raab et al., "Magnetic Position and Orientation Tracking System," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-15, no. 5, 1979, pp. 709-718.
9. G. Welch et al., "High-Performance Wide-Area Optical Tracking: The HiBall Tracking System," *Presence: Teleoperators and Virtual Environments*, vol. 10, no. 1, Feb. 2001, pp. 1-21.
10. E. Foxlin, M. Harrington, and G. Pfeifer, "Constellation: A Wide-Range Wireless Motion-Tracking System for Augmented Reality and Virtual Set Applications," *Proc. Ann. Conf. Computer Graphics and Interactive Techniques (Proc. Siggraph 98),* M.F. Cohen, ed., ACM Press, New York, 1998, pp. 371-378.
11. R. Azuma et al., "Recent Advances in Augmented Reality," *IEEE Computer Graphics and Applications*, vol. 21, no. 6, Nov./Dec. 2001, pp. 34-47.
12. R.T. Azuma, "A Survey of Augmented Reality," *Presence: Teleoperators and Virtual Environments,* vol. 6, 1997, pp. 355-385.
13. S. Feiner. "Augmented Reality: A New Way of Seeing," *Scientific American*, vol. 286, no. 4, Apr. 2002.
14. F. Hohl et al., "Next Century Challenges: Nexus—An Open Global Infrastructure for Spatial-Aware Applications," *Proc. Int'l Conf. Mobile Computing and Networking* (MobiCom 99), T. Imielinski and M. Steenstrup, eds., ACM Press, New York, 1999, pp. 249-255.
15. M. Bajura and U. Neumann, "Dynamic Registration Correction in Video-Based Augmented Reality Systems," *IEEE Computer Graphics and Applications,* vol. 15, no. 5, Sept. 1995, pp. 52-61.
16. G. Welch, *Hybrid Self-Tracker: An Inertial/Optical Hybrid Three-Dimensional Tracking System,* tech. report TR95-048, Univ. of North Carolina at Chapel Hill, Dept. Computer Science, Chapel Hill, N.C., 1995.
17. N. Vallidis, *WHISPER: A Spread Spectrum Approach to Occlusion in Acoustic Tracking*, doctoral thesis, Univ. of North Carolina at Chapel Hill, Dept. Computer Science, Chapel Hill, N.C., 2002.
18. K. Cornelis et al., "Tracking Based Structure and Motion Recovery for Augmented Video Productions," *Proc. ACM Symp. Virtual Reality Software and Tech.*, Addison-Wesley, Reading, Mass., 2001, pp. 17-24.



Greg Welch *is a research associate professor in the Computer Science Department at the University of North Carolina at Chapel Hill. His research interests include virtual/augmented environment tracking systems and telecollaboration. He received a BS in electrical technology from Purdue University and an MS and PhD in computer science from the University of North Carolina at Chapel Hill. He is a member of the IEEE Computer Society and the ACM.*

Eric Foxlin *is founder and chief technology officer of InterSense in Burlington, Massachusetts. His research interests include motion tracking, inertial navigation, sensor fusion, and environment mapping. He received a BA in physics from Harvard and an MS in electrical engineering and computer science from MIT. He is a member of the IEEE Computer Society and Robotics and Automation Society, the ACM, and the Institute of Navigation.*

Readers may contact Welch at University of North Carolina at Chapel Hill, Department of Computer Science, CB# 3175 Sitterson Hall, Chapel Hill, NC 27599-3175; welch@cs.unc.edu; http://www.cs.unc.edu/~welch.

For more information on this or any other computing topic, please visit our Digital Library at http://computer.org/publications/dlib.